IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BOARD OF PATENT APPEALS AND INTERFERENCES

In re patent application of:
Rajput, et al.

Atty. Docket No.: JP920030180US1

Serial No.: 10/727,886

Group Art Unit: 2609

Filed: December 4, 2003

Examiner: Michael C. Colucci

For: LANGUAGE MODELLING FOR MIXED LANGUAGE EXPRESSIONS

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPELLANTS' APPEAL BRIEF

Sirs:

Appellant respectfully appeals the final rejection of claims 1-21, in the Office Action dated August 8, 2007. A Notice of Appeal was timely filed on November 6, 2007.

I. REAL PARTY IN INTEREST

The real party in interest is International Business Machines Corporation, Armonk, New York, assignee of 100% interest of the above-referenced patent application.

II. RELATED APPEALS AND INTERFERENCES

There are no other appeals or interferences known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

III. STATUS OF CLAIMS

Claims 1-21 are all the claims pending in the application and under appeal. Claims 1, 5-9, 13-15, and 19-21 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Poznanski, et al. (U.S. Patent No. 6,397,174), hereinafter referred to as Poznanski, in view of Odell (U.S. Patent No. 6,668,243). Claims 2-3, 10-11, and 16-17 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Poznanski, in view of Odell, in further view of Bahl, et al. (U.S. Patent No. 4,759,068), hereinafter referred to as Bahl. Claims 4, 12, and 18 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Poznanski, in view of Odell, in further view of Lee, et al. (U.S. Patent No. 7,165,019), hereinafter referred to as Lee. None of the claims are allowed; all of the rejections are appealed.

IV. STATUS OF AMENDMENTS

In response to the Office Action mailed August 8, 2007 (referred to herein as the "Office Action"), Appellants filed an after-final Amendment on October 3, 2007. An Advisory Action mailed on October 10, 2007 indicated that the claim amendments of Appellants' after-final Amendment would not be entered. The claims shown in the appendix are shown in their form at the time of the mailing of the Office Action (prior to the October 3, 2007 Amendment).

V. SUMMARY OF CLAIMED SUBJECT MATTER

One feature of the invention is a method for language modelling of mixed language expressions. Claim 1 defines this feature as follows: "A method for language modelling of mixed language expressions." This feature is described at various points in the specification, for example paragraph [0030] describes this feature as follows: "represents a system architecture 100 for language modelling of mixed language expressions." This is shown in Figure 1.

Another feature of the invention is storing word equivalence probabilities relating to words of a first language and words in at least one other language. Claim 1 defines this feature as follows: "storing word equivalence probabilities relating to words of a first language and words in at least one other language." This feature is described at various points in the specification, for example paragraph [0030] describes this feature as follows: "Word equivalence probabilities 130 are also generated and stored for later use." This is shown in Figure 1.

Another feature of the invention is generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein the mixed language word history comprises words in said first language and words in the at least one other language, and wherein the mixed language word history and the monolingual word history each comprise a history of previous words in a sentence-based word sequence. Claim 1 defines this feature as follows: "generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein said mixed language word history comprises words in said first language and words in said at least one other language, and wherein said mixed language word history and said monolingual word history each comprise a history of previous words in a sentence-based word sequence." This feature is described at various points in the specification, for example paragraph [0031] describes this feature as follows: "Finally, a hypothesis is generated for the next word in the base language using monolingual techniques in step

240. Word equivalence probabilities are consulted as required." This is shown in Figure 2.

Another feature of the invention is generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein the monolingual next word hypothesis probabilities predict a next word in the word sequence. Claim 1 defines this feature as follows: "generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein said monolingual next word hypothesis probabilities predict a next word in said word sequence." This feature is described at various points in the specification, for example paragraph [0037] describes this feature as follows: "A language model works on the basis of a given word history. The model attempts to predict the next word in the sequence, given a word sequence history." This is shown in Figure 2.

Another feature of the invention is determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein said probability of said next word predicts a next word in the mixed language expression. Claim 1 defines this feature as follows: "determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein said probability of said next word predicts a next word in said mixed language expression." This feature is described at various points in the specification, for example paragraph [0043] describes this feature as follows: "The next word hypothesis is generated in any of the two languages, base or foreign. The history can be either in the base language or in the foreign language, or in a language that contains words that are a mix of the base and foreign language." This is shown in Figure 2.

Another feature of the invention is a computer program product for language modelling of mixed language expressions, the computer program product comprising computer software recorded on a computer-readable medium for performing. Claim 8 defines this feature as follows: "A computer program product for language modelling of

mixed language expressions, the computer program product comprising computer software recorded on a computer-readable medium for performing." This feature is described at various points in the specification, for example paragraph [0030] describes this feature as follows: "represents a system architecture 100 for language modelling of mixed language expressions." This is shown in Figure 1.

Another feature of the invention is storing word equivalence probabilities relating to words of a first language and words in at least one other language. Claim 8 defines this feature as follows: "storing word equivalence probabilities relating to words of a first language and words in at least one other language." This feature is described at various points in the specification, for example paragraph [0030] describes this feature as follows: "Word equivalence probabilities 130 are also generated and stored for later use." This is shown in Figure 1.

Another feature of the invention is generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein the mixed language word history comprises words in the first language and words in the at least one other language, and wherein the mixed language word history and the monolingual word history each comprise a history of previous words in a sentence-based word sequence. Claim 8 defines this feature as follows: "generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein said mixed language word history comprises words in said first language and words in said at least one other language, and wherein said mixed language word history and said monolingual word history each comprise a history of previous words in a sentence-based word sequence." This feature is described at various points in the specification, for example paragraph [0031] describes this feature as follows: "Finally, a hypothesis is generated for the next word in the base language using monolingual techniques in step 240. Word equivalence probabilities are consulted as required." This is shown in Figure 2.

Another feature of the invention is generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein the monolingual next word hypothesis probabilities predict a next word in the word sequence. Claim 8 defines this feature as follows: "generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein said monolingual next word hypothesis probabilities predict a next word in said word sequence." This feature is described at various points in the specification, for example paragraph [0037] describes this feature as follows: "A language model works on the basis of a given word history. The model attempts to predict the next word in the sequence, given a word sequence history." This is shown in Figure 2.

Another feature of the invention is determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein the probability of teh next word predicts a next word in the mixed language expression. Claim 8 defines this feature as follows: "determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein said probability of said next word predicts a next word in said mixed language expression." This feature is described at various points in the specification, for example paragraph [0043] describes this feature as follows: "The next word hypothesis is generated in any of the two languages, base or foreign. The history can be either in the base language or in the foreign language, or in a language that contains words that are a mix of the base and foreign language." This is shown in Figure 2.

Another feature of the invention is a computer system for language modelling of mixed language expressions. Claim 9 defines this feature as follows: "A computer system for language modelling of mixed language expressions." This feature is described at various points in the specification, for example paragraph [0030] describes this feature as follows: "represents a system architecture 100 for language modelling of mixed language expressions." This is shown in Figure 1.

Another feature of the invention is computer software code means for storing word equivalence probabilities relating to words of a first language and words in at least one other language. Claim 9 defines this feature as follows: "computer software code means for storing word equivalence probabilities relating to words of a first language and words in at least one other language." This feature is described at various points in the specification, for example paragraph [0030] describes this feature as follows: "Word equivalence probabilities 130 are also generated and stored for later use." This is shown in Figure 1.

Another feature of the invention is computer software code means for generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein the mixed language word history comprises words in the first language and words in the at least one other language, and wherein the mixed language word history and the monolingual word history each comprise a history of previous words in a sentence-based word sequence. Claim 9 defines this feature as follows: "computer software code means for generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein said mixed language word history comprises words in said first language and words in said at least one other language, and wherein said mixed language word history and said monolingual word history each comprise a history of previous words in a sentence-based word sequence." This feature is described at various points in the specification, for example paragraph [0031] describes this feature as follows: "Finally, a hypothesis is generated for the next word in the base language using monolingual techniques in step 240. Word equivalence probabilities are consulted as required." This is shown in Figure 2.

Another feature of the invention is computer software code means for generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein the monolingual next word hypothesis probabilities predict a next word in the word sequence. Claim 9 defines this feature as follows: "computer software code means for generating monolingual next word hypothesis

probabilities in the first language based upon the monolingual word history, wherein said monolingual next word hypothesis probabilities predict a next word in said word sequence." This feature is described at various points in the specification, for example paragraph [0037] describes this feature as follows: "A language model works on the basis of a given word history. The model attempts to predict the next word in the sequence, given a word sequence history." This is shown in Figure 2.

Another feature of the invention is computer software code means for determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein the probability of the next word predicts a next word in the mixed language expression. Claim 9 defines this feature as follows: "computer software code means for determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein said probability of said next word predicts a next word in said mixed language expression." This feature is described at various points in the specification, for example paragraph [0043] describes this feature as follows: "The next word hypothesis is generated in any of the two languages, base or foreign. The history can be either in the base language or in the foreign language, or in a language that contains words that are a mix of the base and foreign language." This is shown in Figure 2.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The issues presented for review is whether claims 1, 5-9, 13-15, and 19-21 are unpatentable under 35 U.S.C. §103(a) by Poznanski, in view of Odell, whether claims 2-3, 10-11, and 16-17 are unpatentable under 35 U.S.C. §103(a) by Poznanski, in view of Odell, in further view of Bahl, and whether claims 4, 12, and 18 are unpatentable under 35 U.S.C. §103(a) by Poznanski, in view of Odell, in further view of Lee.

VII. ARGUMENT

A. The Rejection Based on Poznanski and Odell

1. The Position in the Office Action

The Office Action states:

In response to applicants amendment to claims 1, 8, and 9:

Claim Rejections -35 USC § 103

2. Claims 1, 5-9, 13-15, 19-21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Poznanski et al (herein after Poz), US 6397174 in view of Odell US 6668243 B1.

Re claim 1, "Mixed language expressions", Poz teaches the processing of text within natural languages as well as the processing of the text by indexing a bilingual or multilingual dictionary, where a dictionary is functionally equivalent to "history" (Poz Col. I line 19-26).

"Storing word equivalency probabilities", Poz teaches the use of a bilingual dictionary containing bilingual equivalencies (Poz Col. 4 line 34-59). Poz also teaches equivalencies based on a unified measure of probability (Poz Col. 6 line 38-45). "Relating words of a first language and words in at least one other language", Poz teaches a multilingual dictionary, which will have a necessary feature to utilize several languages (Poz Col. 1 line 19-26).

With respect to the first amendment within the current claim 1:

"Monolingual word history" based on a "mixed language word history", Poz teaches a dictionary composed of monolingual or multilingual text (Poz Col. 1 line 19-26). "Stored word equivalency probabilities", Poz teaches the use of a bilingual dictionary containing bilingual equivalencies (Poz Col. 4 line 34-59).

The amended limitation is construed as a mixed language history where the monolingual and mixed history each contain a history of previous words in a sentence based word sequence. Poz fails to teach a previous word history within the dictionaries that are composed of a sentence based word sequence. Odell teaches a history entry for each history where there are predicted words where the history entries are stored in a hash table according to the predicted word identifier (Odell col. 9 line 24- 34). 'sentence based word sequence", Odell also teaches single word recognition extended to the case of

sentences by allowing connections from the end of words to the start of other word so that language model probabilities, which are based upon the likelihood of one word being adjacent to another (Odell col. 1 line 28-39). Therefore, the combined teaching of Poz and Odell would have rendered obvious the generation of word histories composed of one of more languages each containing a history of words.

With respect to the second amendment within the current claim 1:

"Monolingual next word hypothesis probabilities", Poz teaches probabilities assigned dependent on the following word within a text, (Poz Col. 6 line 14-17). Poz also teaches a dictionary composed of monolingual or multilingual text (Poz Col. 1 line 19-26).

"Monolingual next word hypothesis probabilities predict a next word in a word sequence", Poz fails to discretely teach of a next word hypothesis probability. Odell teaches speech recognition systems usually require the calculation of likelihoods which must be computed to compare individual word hypotheses and determine the most likely word (Odell col. 1 line 63 — col. 2 line 14). Therefore, the combined teaching of Poz and Odell would have rendered obvious a monolingual next word hypothesis probability predicting a next word in a word sequence.

With respect to the third amendment within the current claim 1:

"Determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities", Poz teaches probabilities assigned dependent on the following word within a text, (Poz Col. 6 line 14-17). Poz also teaches a dictionary composed of monolingual or multilingual text (Poz Col. 1 line 19-26).

"Probability of said next word predicts a next word in said mixed language expression", Poz fails to teach this limitation based upon the next word hypothesis probability. Odell teaches speech recognition systems that usually require the calculation of likelihoods, which must be computed to compare individual word hypotheses and determine the most likely word (Odell col. 1 line 63 — col. 2 line 14). Odell also teaches language model probabilities as part of a composite model applied to inter-word

connections (Odell col. I line 28-39). Therefore, the combined teaching of Poz and Odell would have rendered obvious a mixed language (multilingual) probability predicting a next word in a mixed language expression.

Claim 5 has been analyzed and rejected with respect to claim 1. A "corpus" is a collection of linguistic speech or text data, as is a dictionary or word history. When read in light of the specification, the term "parallel" is in reference to equivalencies between two languages.

Claim 6 has been analyzed and rejected with respect to claim 1. Poz teaches of translation from a first language to a second language (such as English to Dutch), where either English or Dutch is a foreign language in relation to one another (Poz Col. 4 line 34-39).

Claim 7 has been analyzed and rejected with respect to claim 5. Claim 5 is more descriptive and further limited than claim 7 as to incorporate the use of probabilities as well as a parallel corpus.

Claims 8 has been analyzed and rejected with respect to claim 1. Claim 8 is the product of the method of claim 1.

Claims 9 has been analyzed and rejected with respect to claim 1. Claim 9 is the product of the method of claim 1.

Claims 13 has been analyzed and rejected with respect to claim 5. Claim 13 is the product of the method of claim 5.

Claims 14 has been analyzed and rejected with respect to claim 6. Claim 14 is the product of the method of claim 6.

Claims 15 has been analyzed and rejected with respect to claim 7. Claim 15 is the product of the method of claim 7.

Claims 19 has been analyzed and rejected with respect to claim 5. Claim 19 is the system of the method of claim 5.

Claims 20 has been analyzed and rejected with respect to claim 6. Claim 20 is the system of the method of claim 6.

Claims 21 has been analyzed and rejected with respect to claim 7. Claim 21 is the system of the method of claim 7.

2. Appellants' Position

a. Independent Claims 1, 8, and 9

Appellants traverse the rejections because the prior art of record fails to teach or suggest the claimed features of "determining a next word in a mixed language expression" as defined in independent Claims 1, 8, and 9. Instead, the prior art references determine a next word in a monolingual expression, i.e., words in a single language.

The claimed invention provides methods, computer systems, and computer program products for language modeling for mixed language expressions. More specifically, as described in paragraph 0014 of Appellants' disclosure, the next word within a sentence can be predicted for *mixed language* expressions. This next word can be of the same language as the text of the previous words, or can be from another language.

As further described in paragraph 0015 of Appellants' disclosure, a database of word equivalence probabilities is used as required by a monolingual language generator. The monolingual language generator uses a mixed-language word history to generate a monolingual word history. The monolingual history is in turn used by a monolingual language model. A resulting next-word hypothesis is used by a next-word language change model, which uses word equivalence probabilities to convert the next word in the monolingual word hypothesis to the next word in the foreign language. An expected mixed-language next word can be provided.

Appellants traverse the rejections because the prior art of record fails to teach or suggest the claimed features of "determining a probability of a next word in a mixed language expression" as defined in independent claims 1, 8, and 9. The Office Action expressly acknowledges that Poz fails to teach the limitation wherein the "[p]robability of said next word predicts a next word in said mixed language expression" (Office Action, p. 5, para. 3). However, the Office Action argues that Poz discloses "a dictionary

composed of monolingual or multilingual text" (Office Action, p. 5, para. 2). Moreover, the Office Action argues that Odell discloses speech recognition systems that determine the next word in a word sequence (Office Action, p. 5, para. 3).

First of all, Appellants submit that Odell does not determine a next word in a mixed language expression (independent claims 1, 8, and 9). Instead, Odell determines a next word in a monolingual expression, i.e., words in a single language. In other words, all of the words in the previous word sequence must be in the same language. Odell cannot predict the next word if the spoken word sequence is a "mixed language expression" (independent claims 1, 8, and 9).

Furthermore, Appellants submit that even if one were to combine the multilingual dictionary (e.g., a French-to-English dictionary) of Poz with the teachings of Odell, the word sequence (that the "next word" is derived from) would still be monolingual. There is not teaching or suggestion within the prior art to take the monolingual word sequence of Odell and translate the word sequence to a mixed language sequence using the multilingual dictionary of Poz. Such a translation would consume time and resources without providing a benefit, i.e., does not help accomplish the goal of predicting the "next word". Because the objective of Odell is to take a given word sequence and predict the next word, using the multilingual dictionary of Poz to go back to the given word sequence and translate it would be "working backwards".

Accordingly, Appellants submit that Odell does not determine a next word in a mixed language expression. Instead, Odell determines a next word in a monolingual expression, i.e., words in a single language. Even if one were to combine the multilingual dictionary of Poz with the teachings of Odell, the word sequence (that the "next word" is derived from) would still be monolingual. Therefore, it is Appellants' position that the prior art of record fails to teach or suggest the claimed features of "determining a probability of a next word in a mixed language expression" as defined in independent claims 1, 8, and 9. In view the foregoing, the Board is respectfully requested to reconsider and withdraw this rejection.

b. Dependent Claims 5-7, 13-15, and 19-21

It is Appellants' position that the proposed combination of Poznanski and Odell does not render obvious independent claims 1, 8, and 9 and similarly does not render obvious dependent claims 5-7, 13-15, and 19-21. In view the foregoing, the Board is respectfully requested to reconsider and withdraw this rejection.

B. The Rejection Based on Poznanski, Odell, and Bahl

1. The Position in the Office Action

The Office Action states:

Claims 2-3, 10-11, 16-17 are rejected under 35 U.S.C. 103(a) as being unpatentable over Poznanski et al (herein after Poz), US 6397174 in view of Odell US 6668243 BI and further in view of Bahl et al (herein after Bahl), US 4759068.

Re claim 2, "Monolingual next word hypothesis probabilities", the combined teaching of Poz and Odell teach probabilities assigned dependent on the following word within a text, (Poz Col. 6 line 14-17). The combined teaching also teaches equivalencies based on a unified measure of probability (Poz Col. 6 line 38-45). However the combined teaching fails to teach "summing products" of the former two probabilities. Bahl teaches the state of a machine dependent on the sum of two probabilities (a) and (b) multiplied together. Bahl also goes as far to incorporate the product of a third probability prior to the summation (Bahl Col. 13 line 56-68). Therefore, the combined teaching of Poz, Odell, and Bahl would have rendered obvious the summation of the product of word equivalency and next word hypothesis probabilities.

Re claim 3, "monolingual next word hypothesis probability is a statistical language model", the combined teaching of Poz and Odell teaches next word hypothesis probabilities. However the combined teaching fails to teach that the probability is a statistical language model. A language model such as a Markov model is a type of statistical model. Bahl teaches a Markov model (Bahl Abstract) for phone machines as well as the storage of statistical probabilities (Bahl Col. 16 line 19-25). Therefore, the combined teaching of Poz, Odell, and Bahl would have rendered obvious the

establishment of a monolingual next word hypothesis probability defined as a statistical language model.

Claims 10 has been analyzed and rejected with respect to claim 2. Claim 10 is the product of the method of claim 2.

Claims 11 has been analyzed and rejected with respect to claim 3. Claim 11 is the product of the method of claim 3.

Claims 16 has been analyzed and rejected with respect to claim 2. Claim 16 is the system of the method of claim 2.

Claims 17 has been analyzed and rejected with respect to claim 3. Claim 17 is the system of the method of claim 3.

2. Appellants' Position

a. Dependent Claims 2-3, 10-11, and 16-17

It is Appellants' position that the proposed combination of Poznanski, Odell, and Bahl does not render obvious independent claims 1, 8, and 9 and similarly does not render obvious dependent claims 2-3, 10-11, and 16-17. Bahl is introduced by the Office Action for the limited purpose of illustrating a method of summing products of word equivalence probabilities with monolingual next word hypothesis probabilities (Office Action, p. 7, item 3). Bahl is also cited by the Office Action for the limited purpose of illustrating a statistical language model (Office Action, pp. 7-8, item 3), and does not teach the features of the independent claims that are discussed in the previous sections. In view the foregoing, the Board is respectfully requested to reconsider and withdraw this rejection.

C. The Rejection Based on Poznanski, Odell, and Lee

1. The Position in the Office Action

The Office Action states:

Claims 4, 12, 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Poznanski et al (herein after Poz), US 6397174 in view of Odell US 6668243 81 and further in view of Lee et al (herein after Lee), US 7165019.

Re claim 4, "converting a mixed language", The combined teaching of Poz and Odell disclose the implementation of a equivalency probability but fail to discretely teach the step of converting a mixed language. Lee teaches the handling of mixed-language input of one form into another type of text (Lee Col. 14 line 31-51). Therefore, the combined teaching of Poz, Odell, and Lee would have rendered obvious the use of a monolingual word equivalence probability applied to the step of converting a mixed language word sequence.

Claims 12 has been analyzed and rejected with respect to claim 4. Claim 12 is the product of the method of claim 4.

Claims 18 has been analyzed and rejected with respect to claim 4. Claim 18 is the system of the method of claim 4.

2. Appellants' Position

a. Dependent Claims 4, 12, and 18

It is Appellants' position that the proposed combination of Poznanski, Odell, and Lee does not render obvious independent claims 1, 8, and 9 and similarly does not render obvious dependent claims 4, 12, and 18. Lee is introduced by the Office Action for the limited purpose of illustrating a method of handling a mixed-language input of one form into another type of text (Office Action, pp. 8-9, item 4) , and does not teach the features of the independent claims that are discussed in the previous sections. In view the foregoing, the Board is respectfully requested to reconsider and withdraw this rejection.

D. CONCLUSION

In view the forgoing, the Board is respectfully requested to reconsider and withdraw the rejections of claims 1-21.

Please charge any deficiencies and credit any overpayments to Attorney's Deposit Account Number 09-0441.

Respectfully submitted,

Date: January 3, 2008

/Frederick W. Gibb, III/
Frederick W. Gibb, III
Registration No. 37,629

Gibb & Rahman, LLC
2568-A Riva Road, Suite 304
Annapolis, MD, 21401
Voice: (410) 573-6501
Fax: (301) 261-8825
Customer No. 29154

IX. CLAIMS APPENDIX

1. A method for language modelling of mixed language expressions, said method comprising the steps of:

storing word equivalence probabilities relating to words of a first language and words in at least one other language;

generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein said mixed language word history comprises words in said first language and words in said at least one other language, and wherein said mixed language word history and said monolingual word history each comprise a history of previous words in a sentence-based word sequence;

generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein said monolingual next word hypothesis probabilities predict a next word in said word sequence; and

determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein said probability of said next word predicts a next word in said mixed language expression.

2. The method as claimed in claim 1, further comprising the step of summing products of word equivalence probabilities with respective monolingual next word hypothesis probabilities.

3. The method as claimed in claim 1, wherein the monolingual next word hypothesis probability is a statistical language model.

4. The method as claimed in claim 1, further comprising the step of converting a mixed language word sequence to a monolingual word sequence using word equivalence probabilities.

5. The method as claimed in claim 1, further comprising the step of determining the word equivalence probabilities based upon a parallel text corpus that has corresponding expressions in the first language and the at least one other language.

6. The method as claimed in claim 1, further comprising the step of determining a probability of a foreign language next word hypothesis given a base language word history.

7. The method as claimed in claim 1, further comprising the step of using a parallel text corpus that has corresponding expressions in the first language and the at least one other language.

8. A computer program product for language modelling of mixed language expressions, the computer program product comprising computer software recorded on a computer-readable medium for performing the steps of:

storing word equivalence probabilities relating to words of a first language and words in at least one other language;

generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein said mixed language word history comprises words in said first language and words in said at least one other language, and wherein said mixed language word history and said monolingual word history each comprise a history of previous words in a sentence-based word sequence;

generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein said monolingual next word hypothesis probabilities predict a next word in said word sequence; and

determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein said probability of said next word predicts a next word in said mixed language expression.

9. A computer system for language modelling of mixed language expressions, the computer system comprising:

computer software code means for storing word equivalence probabilities relating to words of a first language and words in at least one other language;

computer software code means for generating a monolingual word history in the first language based upon a mixed language word history and using the stored word equivalence probabilities, wherein said mixed language word history comprises words in said first language and words in said at least one other language, and wherein said mixed language word history and said monolingual word history each comprise a history of previous words in a sentence-based word sequence;

computer software code means for generating monolingual next word hypothesis probabilities in the first language based upon the monolingual word history, wherein said monolingual next word hypothesis probabilities predict a next word in said word sequence; and

computer software code means for determining a probability of a next word in a mixed language expression based upon the monolingual next word hypothesis probabilities and the stored word equivalence probabilities, wherein said probability of said next word predicts a next word in said mixed language expression.

10. The computer program product as claimed in claim 8, further comprising the step of summing products of word equivalence probabilities with respective monolingual next word hypothesis probabilities.

11. The computer program product as claimed in claim 8, wherein the monolingual next word hypothesis probability is a statistical language model.

12. The computer program product as claimed in claim 8, further comprising the step of converting a mixed language word sequence to a monolingual word sequence using word equivalence probabilities.

13. The computer program product as claimed in claim 8, further comprising the step of determining the word equivalence probabilities based upon a parallel text corpus that has corresponding expressions in the first language and the at least one other language.

14. The computer program product as claimed in claim 8, further comprising the step of determining a probability of a foreign language next word hypothesis given a base language word history.

15. The computer program product as claimed in claim 8, further comprising the step of using a parallel text corpus that has corresponding expressions in the first language and the at least one other language.

16. The computer system as claimed in claim 9, further comprising computer software code means for summing products of word equivalence probabilities with respective monolingual next word hypothesis probabilities.

17. The computer system as claimed in claim 9, wherein the monolingual next word hypothesis probability is a statistical language model.

18. The computer system as claimed in claim 9, further comprising computer software code means for converting a mixed language word sequence to a monolingual word sequence using word equivalence probabilities.

19. The computer system as claimed in claim 9, further comprising computer software code means for determining the word equivalence probabilities based upon a parallel text corpus that has corresponding expressions in the first language and the at least one other language.

20. The computer system as claimed in claim 9, further comprising computer software code means for determining a probability of a foreign language next word hypothesis given a base language word history.

21. The computer system as claimed in claim 9, further comprising computer software code means for using a parallel text corpus that has corresponding expressions in the first language and the at least one other language.

X. EVIDENCE APPENDIX

There is no other evidence known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

XI. RELATED PROCEEDINGS APPENDIX

There is no other related proceedings known to Appellants, Appellants' legal representative or Assignee which would directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.